

Uncovering Expertise-Related Differences in Troubleshooting Performance: Combining Eye Movement and Concurrent Verbal Protocol Data

Citation for published version (APA):

Van Gog, T., Paas, G. W. C., & van Merriënboer, J. J. G. (2005). Uncovering Expertise-Related Differences in Troubleshooting Performance: Combining Eye Movement and Concurrent Verbal Protocol Data. *Applied Cognitive Psychology*, 19(2), 205-221. <https://doi.org/10.1002/acp.1112>

DOI:

[10.1002/acp.1112](https://doi.org/10.1002/acp.1112)

Document status and date:

Published: 02/03/2005

Document Version:

Peer reviewed version

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 02 Jul. 2022

Open Universiteit
www.ou.nl



Running head: UNCOVERING EXPERTISE-RELATED DIFFERENCES

This is a pre-print of the article that was published as

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205-221.

Full text available from <http://dx.doi.org/10.1002/acp.1112>

© 2005 John Wiley & Sons, Ltd.

Uncovering Expertise-Related Differences in Troubleshooting Performance: Combining Eye
Movement and Concurrent Verbal Protocol Data

Tamara van Gog, Fred Paas, and Jeroen J. G. van Merriënboer

Educational Technology Expertise Center, Open University of the Netherlands, Heerlen

Keywords: verbal protocols; eye tracking; troubleshooting; expertise

Abstract

This study explored the value of eye movement data for uncovering relatively small expertise-related differences in electrical circuit-troubleshooting performance, and describes that value in relation to concurrent verbal protocols. Results show that in the '*problem orientation*' phase, higher expertise participants spent relatively more time, had a shorter mean fixation duration, and fixated more on a major fault-related component than lower expertise participants. In the '*problem formulation*' part of the '*problem formulation and action decision*' phase, the mean fixation duration of the higher expertise participants was longer. In the '*action evaluation and next action decision*' phase, higher expertise participants spent relatively more time than the lower expertise participants. Over the different phases, only the mean fixation duration of the higher expertise participants differed significantly. The relation between the eye movement and concurrent verbal protocol data is qualitatively described. The results are discussed in perspective of the combined value of eye tracking and concurrent reports for expertise research and instructional design.

Uncovering Expertise-Related Differences in Troubleshooting Performance: Combining Eye Movement and Concurrent Verbal Protocol Data

Technical troubleshooting, that is, diagnosing and repairing faults in a technical system, is considered a complex process to carry out and learn. Effective performance of a troubleshooting task requires adequate domain, system, and strategic knowledge (organized in mental models), and adequate reasoning based on this knowledge (Gitomer, 1988; Schaafstal, Schraagen, & Van Berlo, 2000). By definition, experts possess much more extensive knowledge than non-experts, and in addition, their knowledge is more effectively organized and better accessible in long-term memory (Chi, Glaser, & Farr, 1988; Ericsson & Lehmann, 1996), which makes their performance more efficient.

For example, expert troubleshooters' system knowledge, allows them to exhibit *failure-based reasoning* when troubleshooting familiar systems. That is, their mental models contain a large body of knowledge of previously encountered failures that caused similar malfunctioning, which allows them to diagnose and act based on that similarity. In the absence of this high amount of experience, non-experts have to rely on *system-based reasoning*, that is, they have to build a mental representation of the system and use that representation to reason about the system's behavior in order to diagnose possible faults. The same goes for experts when troubleshooting unfamiliar systems, however, in those cases the amount and organization of their domain knowledge allow them to build a representation faster, and their strategic knowledge allows them to apply effective strategies (e.g., the structured approach to troubleshooting described by Schaafstal et al., 2000). Non-experts have to rely on weaker and domain-general strategies (e.g., means-ends analysis, which relies on backward reasoning; Patel, Arocha, & Kaufman, 1994; Patel, Groen, & Norman, 1993; Sweller, 1988). Evidence for the fact that the amount and organization of experts' domain knowledge support a faster construction of mental system representations comes from the work of Egan and Schwartz (1979). They found that more skilled technicians formed chunks of components in schematic drawings according to their functional units, and were able to recall more information than less skilled technicians after brief

exposure to a drawing. In addition, chunking mechanisms (Gobet et al., 2001) reduce working memory load so that experts can devote more cognitive capacity to reasoning. Approximately 5-9 elements or chunks of information can be held in working memory simultaneously (and less when information is not only to be remembered, but also processed; Sweller, 2004). So, when the chunks that are formed or retrieved from long-term memory are larger (e.g., an entire functional unit instead of one component), more information can be held in working memory. Hence, an individual with more expertise is able to keep the same amount of information in working memory while using less capacity than an individual with less expertise, and can therefore devote more cognitive capacity to reasoning.

Although an extensive body of research exists on expert-novice differences in knowledge, memory, and performance in a substantial number of domains, among which, technical troubleshooting (e.g., in general: Chi et al., 1988; Ericsson & Lehmann, 1996; in troubleshooting: Gitomer, 1988; Schaafstal et al., 2000), much less is known about the relatively long intermediate phase on the developmental continuum from being a novice to becoming an expert in a domain.¹ For most complex cognitive skills, this intermediate phase in which students gradually acquire competence, can have a very long duration (cf. “the 10-year rule of necessary preparation” for attaining excellence; Ericsson & Lehmann, 1996). Consequently, there are expertise differences, or (sub)levels of less to more skilled performance, *within* this phase. Research on the differences in knowledge and performance at those sublevels of expertise will advance the insights of expertise research. Moreover, such research on more subtle expertise differences is imperative for instructional design (Alexander, 2003; Van Gog, Ericsson, Rikers, & Paas, in press). Instructional designers acknowledge that in order to foster students’ expertise development as far as possible during a formal instructional period, instruction should be adaptive to the individual learner’s level of expertise (Shute & Towle, 2003). However, educational research has not systematically addressed the questions of exactly how expertise in a domain develops, and what aspects of performance distinguish students at different (sub)levels of expertise (Alexander, 2003; Van Gog et al., in press).

Uncovering Expertise-Related Performance Differences: Process-Tracing Techniques

For studying expertise-related differences in performance on complex cognitive tasks, process-tracing techniques are very promising because “... the data that are recorded are of a pre-specified type (e.g., verbal reports, eye movements, actions) and are *used to make inferences about the cognitive processes or knowledge underlying task performance*” (Cooke, 1994, p. 814; italics added). Verbal reports, such as concurrent (‘think aloud’) and retrospective reporting (Ericsson & Simon, 1993) are probably the most widely used process-tracing techniques. With the method of concurrent reporting, participants are instructed to “think aloud”, that is, verbalize everything that comes to mind, *while* they are working on a task. With the method of retrospective reporting, participants are instructed to verbalize the thoughts they had during problem solving *immediately after* finishing the task. Both methods can result in verbal protocols that allow for making inferences about cognitive processes, but to ensure validity of those inferences, the wording of verbalization instructions and prompts is crucial (Ericsson & Simon, 1993). Furthermore, their on-line (concurrent) and off-line (retrospective) generation may lead to differences in the kind of information contained in the protocols, and hence, the kind of inferences made (for an in-depth discussion of these methods, instructions, and results, see Ericsson & Simon, 1993).

Eye tracking, that is, recording eye movement data while participants are working on a task, is less commonly used as a process-tracing method. However, eye movement data provide insight in the allocation of attention and therefore allow –albeit cautious- inferences to be made about cognitive processes (Rayner, 1998). Attention can shift in response to exogenous or endogenous cues (Rayner, 1998; Stelmach, Campsall, & Herdman, 1997). Whereas exogenous shifts occur mainly in response to salient features in the environment, endogenous shifts are driven by knowledge of the task, of the environment, and of the importance of information sources, and are therefore influenced by expertise (cf. Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003).² For example, Haider and Frensch’s (1999) information-reduction hypothesis, stating that with practice, people learn to ignore task-redundant

information and limit their processing to task-relevant information, was corroborated by eye movement data, and in the domain of chess it was found that experts fixated proportionally more on relevant pieces than non-expert players (Charness, Reingold, Pomplun, & Stampe, 2001). Furthermore, eye movement data can provide information about the cognitive load particular cognitive processes impose. For example, pupil dilation (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004), and fixation duration (Underwood, Jebbett, & Roberts, 2004) are known to increase with increased processing demands (task difficulty), whereas the length of saccades decreases (for an in-depth discussion of eye movement data and cognitive processes, see Rayner, 1998).

So, eye movement data presumably have the potential to show differences in the problem-solving process at a more fine-grained level than verbal protocol data, as well as to provide information about the cognitive demands of those processes that cannot be inferred from verbal protocols. Therefore, the combination of eye movement and concurrent verbal protocol analysis to obtain insight into the content of ongoing cognitive processes and the cognitive demands they impose may be especially useful when investigating relatively small expertise differences.

This study aims to explore the value of eye movement data and the combination of these data with concurrent verbal protocol data for discovering expertise-related differences in troubleshooting performance between students at lower and higher sublevels of expertise in the early intermediate phase. In secondary education science curricula, simple (and nowadays often computer-simulated) troubleshooting tasks are used to teach and test students' ability to not only memorize certain principles (e.g., Ohm's law), but also understand their working and use this to reason about a technical systems' behavior. The troubleshooting tasks used in this study consist of malfunctioning computer-simulated electrical circuits. In the process of solving such problems, it is possible to distinguish the following phases based on the physical (i.e., mouse/keyboard) actions taken on the circuit:

- *Problem orientation*: orienting to the circuit;

- *Problem formulation and action decision*: formulating a problem description and deciding on the first action, which involves diagnosis of possible causes;
- *Action evaluation and next action decision*: evaluating the outcome of the action and deciding on the next, which again involves diagnosis; and
- *Evaluation*: a final evaluation of the problem solution (e.g., in terms of outcome versus costs).

Because the diagnosed and tested possible cause may not be the actual or only cause of the malfunctioning, troubleshooting is often a cyclic process (even for individuals with a substantial amount of expertise), and therefore the third phase will be repeated until the problem is solved and the final solution can be evaluated. Note that in phase 2 and 3, two cognitive processes presumably take place ('subphases'), although these cannot be distinguished as separate phases based on the physical (i.e., mouse/keyboard) actions taken on the circuit.

This study centers on the first three phases of the process, '*problem orientation*', '*problem formulation and action decision*' (first action), and '*action evaluation and next action decision*' (second action). Given that all participants are in the early intermediate phase, we expect the higher expertise participants to spend more time on problem orientation because they will try to build a mental representation of the problem, as well as on problem formulation, because they can use their mental models in combination with the representation to reason about how this specific circuit should function if it was intact. In contrast, lower expertise participants will find it harder to build a representation and be more likely to test the functioning of the circuit and try to use this information to generate hypotheses.³ Higher expertise participants might spend relatively more time on deciding on actions and evaluating them, because they might try to consider the impact of their action and will evaluate (or monitor) whether it has gotten them closer to their goal, whereas lower expertise participants would be more likely to try what comes to mind and use the outcome to generate new hypotheses, without a clear sense of the goal. This refers back to the problem orientation and formulation: if they do not consider how the circuit should function if it was intact, they have not explicitly and concretely defined the

goal state, which makes monitoring difficult. With regard to processing demands as reflected in fixation duration, we expect orientation and evaluation to be less cognitively demanding than reasoning, that is, formulating the problem and deciding on actions, and we expect all these processes to be more demanding for lower than higher expertise participants. In line with the findings by Haider and Frensch (1999) and Charness et al. (2001) we hypothesize that in the ‘*problem orientation*’ phase, higher expertise participants will have a higher proportion (percentage) of fixations on and gaze switches between components related to major faults in the electrical circuit. Finally, we will qualitatively describe the relation between the eye movement and concurrent verbal protocol data, to get an indication of the unique contribution eye movement data may make to the investigation of cognitive processes.

Method

The data reported here were collected as part of a larger experiment. The first aim of this experiment was to compare three verbal methods of knowledge elicitation, *concurrent reporting*, *retrospective reporting* and *cued retrospective reporting* based on a record of eye movements and mouse/keyboard operations, on the types of problem-solving information they elicited and to study the possible influence of expertise on the results. The second aim was the one reported here: to investigate the value of eye movement data for uncovering expertise-related performance differences and to describe that value in relation to concurrent verbal protocols. These data were obtained in the fourth condition included in the experiment: *concurrent reporting with eye tracking*.

Design

The experiment was set up as a within-subjects, balanced Latin square design, resulting in four sequences of reporting methods and tasks, to which participants were randomly assigned (e.g., the first sequence was: concurrent reporting, tasks 1 and 2; concurrent reporting with eye tracking, tasks 3 and 4; retrospective reporting, tasks 5 and 6; cued retrospective reporting, tasks 7 and 8). The task reported on here was the first of the concurrent reporting with eye tracking condition. It was the same for each participant, irrespective of their assigned sequence, although

it could be either the first, third, fifth, or seventh task they worked on.

Participants

Participants were students in their fifth year of pre-university education or in their first or second year of higher professional education. All participants had uncorrected good eyesight or good eyesight when corrected with hydrophilic contact lenses, and all had studied *at least* the basic theory of electricity (i.e., all of them knew the relevant domain principles and the function of the circuit components). Participation was voluntary and was rewarded with € 12.50 after the experiment. From the entire group of 26 participants, the 5 participants with the highest and lowest expertise were selected for this study, based on a measure of performance efficiency on all tasks in the experiment. This measure consists of a combination of performance, mental effort, and time-on-task scores (see Table 1 for these scores and the ‘apparatus and materials’ section for a description of the measures), and is based on the rationale that students of higher expertise are able to obtain equal or higher performance with lower investment of time and mental effort (Anderson & Fincham, 1994; Kalyuga & Sweller, in press; see also Yeo & Neal, 2004).⁴

<...Insert Table 1 about here...>

So, the ‘higher expertise’ group consisted of the 5 participants with the highest performance efficiency. All of them were in the first or second year of higher professional education. One of them had worked on sequence 1, two had worked on sequence 2, one on sequence 3, and one on sequence 4. The ‘lower expertise’ group consisted of the 5 participants with the lowest performance efficiency. Four of them were in the fifth year of pre-university education, and 1 was in the first year of higher professional education. Two of them had worked on sequence 1, one had worked on sequence 2, one on sequence 3, and one on sequence 4.

Apparatus and Materials

Troubleshooting task. All troubleshooting tasks in the experiment consisted of malfunctioning electrical circuits, constructed and offered in a simulation program, Crocodile Physics[®], version 1.5. A science teacher constructed them, at the level of difficulty of fourth year

higher general secondary education and pre-university education. All circuits contained at least the following components: a toggle switch, a lamp, a battery, a voltmeter and an ammeter. Those components were supplemented in varying ways by the following components: toggle switches, push switches, lamps, batteries, voltmeters, ammeters, resistors, variable resistors, fuses, buzzers, and gears driven by constant speed motors.

Each of the circuits contained multiple faults. For example, components (like batteries and meters) could be connected in the wrong way, components could be short-circuited, and there could be problems with the voltage or current, resulting from too high or too low power supplies (batteries) or resistance (resistors). Participants were instructed to troubleshoot the circuits so that they would function properly: a) all components with outwardly observable functions should function visibly when the circuit was closed (e.g., lamps should burn and gears should turn visibly), b) the repaired circuit should contain at least the same components as the initial, malfunctioning circuit, that is, components could be added, or changed, but not removed, c) all components should be properly connected, and d) in the case of multiple switches, the circuit should function when all switches were in the “on” position.

In the introduction preceding the tasks, participants were acquainted with the functioning of Crocodile Physics[®] and with thinking aloud by means of a simple practice task. No domain-specific information was provided either in the introduction or with the tasks.

The first troubleshooting task of the concurrent reporting with eye tracking condition is shown in Figure 1.

<...Insert Figure 1 about here...>

The first major fault in this circuit is that the voltage of the battery (3V) is far too low to provide enough power for three lamps (which have a maximum rating of 9V). Hence, when the functioning of the circuit is tried by pressing the switch, participants will see that the lights do not burn and can read the low current from the ammeter. A second major fault is that lamp 1 is short-circuited by the switch, so that even if the battery voltage is raised, this lamp will only burn when the switch is *open*. A minor fault is that the meters are connected in the wrong direction (as

indicated by the ‘-’ symbol). To repair the circuit (in correspondence with the definition of proper functioning), the meters should be connected in the right direction, the switch and lamp 1 should be connected in series, and the voltage of the battery should be raised to the range of 15V to 22V to make all lights burn visibly (or alternatively, the resistance could be reduced by lowering the value of the resistor and raising the voltage within a lower range –e.g., in the case of setting the resistor to 10 Ohm, the voltage should be raised to the range of 12V to 14V). So, the task allows multiple approaches to reaching a correct solution, by allowing the actions to be carried out in a different order, and by allowing choice between different options to reach the same goal.

Performance rating. Participants’ task solutions were scored on the following aspects: a) functioning at each switch position, b) intensity (e.g., of the lamps, of motor rotation), c) optimal functioning (e.g., in the case of a variable resistor and a lamp, the resistor really had to work as a dimmer for the lamp, not just as a resistor), d) proper connection of components, e) proper direction of meters, f) no unnecessary addition of components, g) no extreme values of components. Not all aspects applied to each task, for example, when the meters were already properly directed, the proper direction of meters was not relevant. Also, for some tasks certain aspects could have a maximum score of two points instead of one, when a circuit had multiple switches for example, two points could be gained for aspect ‘a’. For each of the tasks, a maximum score of 7 points could be gained. To determine the reliability of the scoring form, two raters scored the performance of 15 participants. The reliability was .94 (Intraclass Correlation Coefficient 3,2; Shrout, & Fleiss, 1979), and the internal consistency was .97 (Cronbach’s alpha).

Mental effort. Participants indicated the amount of mental effort it took them to complete the task on a 9-point rating scale ranging from 1 “very, very low effort” to 9 “very, very high effort”, which is known to be a reliable measure of experienced mental effort (Paas, 1992; Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

Registration of eye movements. Participants’ eye movements were recorded with a 50Hz

video-based remote eye-tracking device (RED) from SensoMotoric Instruments (SMI) with an angular resolution of less than 0.5° . This infrared camera was placed under the 21-inch screen of the stimulus PC, located in a recording room. The resolution of the stimulus PC's screen was set at 1024 x 768 pixels. The size of the diagram on the screen was 740 x 480 pixels. An adjustable forehead rest was placed in front of the screen, so that the participant's eyes were positioned at a distance of approximately 70 cm from the center of the screen. On a PC in an adjoining observation room, I-View software (SMI) operated the camera and the calibration of the eye-tracking system. An extra mouse, keyboard and monitor were connected to the stimulus PC and located in the observation room. This enabled the experimenter to perform the necessary actions on the stimulus PC when calibrating the system from the observation room. GazeTracker™ software (Lankford, 2000) ran on the stimulus PC to register participants' eye movements and their mouse and keyboard operations. These registration files also enabled determination of time-on-task. The recording room was visible from the observation room through a one-way screen, and microphones that were attached to a digital audio-recorder enabled verbal communication between both rooms and the recording of participants' verbalizations.

Procedure

Before they started working on the task, the eye-tracking system had to be calibrated, so the experimenter asked the participants to place their head in the forehead rest and instructed the participants: "I am going to record your eye movements while you are working on the next task, so I will calibrate the eye-tracking system before you start. In a minute you will see a red square appearing on your screen, please follow it with your eyes". After calibration, the experimenter instructed the participants "Thank you, the system is calibrated. Please think aloud while you are working on the next task". The GazeTracker™ software was started to record participants' eye movements and mouse/keyboard operations. Participants' verbalizations were recorded on digital audiocassettes (and were transcribed after the session). After they finished the task participants went to the next screen where they indicated their perceived mental effort on the 9-point rating scale, and then waited for instructions from the experimenter.

Data Reduction and Analysis

GazeTracker™ saves the eye movement and mouse/keyboard data in a Microsoft Access database. Based on the mouse-click data in participants' databases, it was determined how much time they spent on each phase. The first phase, '*problem orientation*', started at 0 s and ended when the participant tried the functioning of the circuit by pressing the switch. The second phase, '*problem formulation and action decision*', started at the end of the first phase and ended when the participant initiated the first action (repair) on the circuit. The third phase, '*action evaluation and next action decision*', started after completion of the first action (so there was a time lag between the end of the second phase and the onset of the third, during which the action was carried out), and ended with the initiation of the second action. The time participants spent on each phase was converted to a percentage of their total time on task.

To be able to roughly distinguish between the *problem formulation* and *action decision* subphases in phase 2, or between the *action evaluation* and *next action decision* subphases in phase 3, these phases were split half-way (end-time minus start-time divided by two), and the mean fixation duration for the first and second half of phases 2 and 3 was calculated.

We used a dispersion-based method of fixation identification (Salvucci, 1999), and identified fixation points by a minimum number of 3 gaze points that fell within a certain dispersion, that is, were grouped within a radius of 40 pixels, and together had a minimal duration of 100 ms. Using GazeTracker™ the fixation data for each phase were exported to Microsoft Excel. This resulted in an overview of the coordinates and duration of each fixation, and –when applied- the LookZone (Area of Interest) in which a fixation fell, as well as “summary” data like the number of fixations, mean fixation duration, possible tracking time lost, percentage of time fixated, etcetera (and where LookZones were applied these “summary” data were also provided for each LookZone). The area around the circuit components that were related to the primary faults, that is, the battery, lamp 1 and the switch, were defined as LookZones. The LookZones were 115 by 146 or 146 by 115 pixels, depending on whether the components were horizontally or vertically located.

Results

Because of the small sample sizes, non-parametric tests were used for the analyses. First, the ‘higher expertise’ and ‘lower expertise’ participants’ relative time spent in a phase and mean fixation duration in a phase (and for phase 1 also the percentage of fixations on the battery and the number of gaze switches between the switch and lamp 1) were compared using Mann-Whitney U Tests. For these analyses, the exact 2-tailed significance is reported. Given our initial hypotheses, a less stringent significance level of .10 is used to avoid type II error (i.e., increase power). Second, ‘lower expertise’ and ‘higher expertise’ participants’ fluctuation of mean fixation duration over time was analyzed using a Friedman Test for K Related Samples with a Nemenyi post-hoc procedure. Third, the concurrent verbal protocol data were qualitatively related to the eye movement data.

Phase 1: Problem Orientation

Table 2 shows the means and standard deviations for the absolute time spent (s), the relative time spent (%), the number of fixations, and fixation duration in each phase. The medians and quartiles for the percentage of time and the mean fixation duration data (used in the non-parametric tests) are summarized in Table 3.

<...Insert Table 2 about here...>

<...Insert Table 3 about here...>

The ‘higher expertise’ participants spent relatively more time on this phase than ‘lower expertise’ participants ($U = 4.0, p = .095$). The mean fixation duration in this phase was higher for the ‘lower expertise’ participants ($U = 4.0, p = .095$).

The median of the ‘higher expertise’ participants’ percentage of fixations on the battery of the total amount of fixations on the diagram⁵ in this phase was 10.53 (interquartile range [IQR] = 5.00 – 15.88), that of the ‘lower expertise’ participants was 5.45 (IQR = .00 - 7.69). The ‘higher expertise’ participants had a higher percentage of fixations on the battery than the ‘lower expertise’ participants ($U = 4.0, p = .087$).

The median of the ‘higher expertise’ participants’ gaze switches between the switch and

lamp 1 was 2.0 (IQR = .5 - 4.0), for the ‘lower expertise’ participants it was 1.0 (IQR = .0 - 1.0). The number of gaze switches of the ‘higher expertise’ participants was not significantly higher, although the difference approached significance ($U = 5.5, p = .175$)

Phase 2: Problem Formulation and Action Decision

The relative time spent on this phase did not differ significantly between the ‘higher expertise’ participants and the ‘lower expertise’ participants ($U = 6.0, p = .222$), and neither did the mean fixation duration in this phase ($U = 6.0, p = .222$).

In the *problem formulation* part of phase 2 (2.1), the mean fixation duration of the ‘higher expertise’ participants was higher than that of the ‘lower expertise’ participants ($U = 4.0, p = .095$). However, in the *action decision* part of phase 2 (2.2), the mean fixation duration did not differ significantly between the ‘higher expertise’ participants and the ‘lower expertise’ participants ($U = 12.0, p = 1.000$).

Phase 3: Action Evaluation and Next Action Decision

The ‘higher expertise’ participants spent relatively more time on this phase than ‘lower expertise’ participants ($U = 3.0, p = .056$). However, the mean fixation duration did not differ between the ‘higher expertise’ participants and the ‘lower expertise’ participants, neither in the entire phase ($U = 6.0, p = .222$), nor in the *action evaluation* ($U = 12.0, p = 1.000$) or *next action decision* ($U = 6.0, p = .222$) parts of this phase.

Mean Fixation Duration over Phases

Figure 2 shows the median values of the ‘lower expertise’ and ‘higher expertise’ participants’ mean fixation duration over time (these values are also shown in Table 3).

<...Insert Figure 2 about here...>

A Friedman Test (2-tailed) showed no significant differences in mean fixation duration over the phases for the ‘lower expertise’ participants ($\chi^2 = 6.40, df = 4, p = .174$). The ‘higher expertise’ participants’ mean fixation duration over the phases differed significantly ($\chi^2 = 10.40, df = 4, p = .022$). A Nemenyi post-hoc procedure shows that the mean fixation duration in phase 1 is significantly lower than that in phase 2.1, 2.2 and 3.2, and that the mean fixation duration in

phase 2.1 is significantly higher than that in phase 3.1.⁶

Combining Eye Movement and Concurrent Verbal Protocol Data

In Table 4, participants' verbalizations in Phase 1, 2 and 3 are reported. Verbalizations of 1 participant in the 'higher expertise' group are missing data due to a recording error.

<...Insert Table 4 about here...>

A qualitative interpretation of the verbal protocol data is in line with the finding that 'higher expertise' participants spent relatively more time on the '*problem orientation*' phase. These data suggest that the 'lower expertise' participants hardly oriented and were immediately focused on testing the functioning of the circuit (e.g., 'lower expertise' participants 3, 4, and 5), whereas the 'higher expertise' participants seemed more inclined to make an inventory of the circuits' components and predict it's functioning before trying it (e.g., 'higher expertise' participants 1 and 3). Probably related to this predictive behavior, is the finding that the 'higher expertise' participants devoted more attention (a higher proportion of fixations) to the battery, which was a major fault-related component. Even though the 'higher expertise' participants tended to engage in more thorough orientation, the results on comparison of mean fixation durations in this phase suggest that this phase led to more extensive processing for the 'lower expertise' participants. Furthermore, the *within-subjects* findings (Figure 2) suggest that for the 'higher expertise' participants this orientation led to less extensive processing than problem formulation and deciding on actions.

It seems that the 'higher expertise' participants stated their problem formulation with somewhat more certainty than the 'lower expertise' participants, which might relate to the finding from the mean fixation duration data that the 'higher expertise' participants showed more extensive processing in the 'problem formulation' part of the '*problem formulation and action decision*' phase.

With regard to the finding that 'higher expertise' participants spent relatively more time on the '*action evaluation and next action decision*' phase, the verbal protocols suggest that this might be because 'lower expertise' participants did not evaluate the outcome of their action in as

much detail as the ‘higher expertise’ participants did. Although the higher expertise participants tended to engage in more thorough action evaluation, there were no significant differences in mean fixation duration between groups. The *within-subjects* findings suggest that the ‘higher expertise’ participants’ action evaluation led to less extensive processing than problem formulation and deciding on actions.

Discussion

Our expectation that the ‘higher expertise’ participants would spend relatively more time on problem orientation, problem formulation, deciding on actions and evaluating them, was confirmed for the *‘problem orientation’* and *‘action evaluation and next action decision’* phases, but not for the *‘problem formulation and action decision’* phase. The verbal data suggest that the ‘lower expertise’ participants hardly oriented at all, but were immediately focused on testing the functioning of the circuit, whereas the ‘higher expertise’ participants tended to show predictive behavior during orientation, which may have helped to limit the relative time they spent on the *‘problem formulation and action decision’* phase. This is in line with our assumption that the mental models of individuals with more expertise are better developed. The lower expertise students should have been capable of making a mental representation of the circuit by inventorying its components, which in turn would have allowed them to reason about the circuits’ behavior without testing it (given that they knew the basic principles and the function of components). However, the fact that they favored the opposite, to test it first and then start reasoning, may have been due to the fact that this imposed less cognitive load for them than constructing a mental representation. Probably, the ‘higher expertise’ participants’ more effective mental models allowed them to allocate capacity to the construction of a representation and even to mentally test the circuit and make predictions about its behavior, as well as to focus on critical components (the higher proportion of fixations on the battery; although the number of gaze switches between the switch and lamp 1 did not differ significantly). Apart from differences in cognitive load, this may also relate to differences in metacognitive knowledge. Individuals with less expertise are known to start working in the first direction that comes to mind, and tend to

stick to that direction, whereas individuals with more expertise are known to spend relatively more time before deciding on a direction and evaluate (or monitor) whether their actions bring them closer to the desired goal (Schoenfeld, 1987).

Even though ‘higher expertise’ participants spent relatively more time on problem orientation (and possibly on action evaluation as well –that cannot be inferred from these data), these phases seemed to impose less processing demands for them than formulating the problem and deciding on actions. A possible explanation might be that shorter fixations might signal perceptual encoding processes, whereas longer fixations might represent problem-solving processes. Unexpectedly, for ‘lower expertise’ participants all these processes seemed to impose equal processing demands. Furthermore, with exception of the ‘*problem orientation*’ phase in which the ‘lower expertise’ participants did have a longer mean fixation duration (as we expected), the ‘higher expertise’ participants tended to have a longer mean fixation duration in all other phases, although this was only significantly longer in the ‘*problem formulation*’ subphase. This is surprising, given that the ‘higher expertise’ participants had a significantly lower self-reported overall mental effort than the ‘lower expertise’ participants. Apparently, the cognitive demands that are measured through the fine-grained fixation data and the more global perceived mental effort scale differ.

Finally, a number of critical observations must be made with regard to this study. First of all, participants had been randomly assigned to task sequences, and all four sequences were represented in both the ‘lower expertise’ and the ‘higher expertise’ groups. Hence, our findings are not likely to be sequence artifacts. Second, a weak point of the use of eye movement (and mouse/keyboard) data is that it does not enable a distinction to be made between the cognitive processes of problem formulation/ action evaluation and action decisions. In order to get some indication, we made a rough distinction by splitting phases 2 and 3 in half, but these results should be interpreted with caution. However, the verbal protocol data did not allow us to make a better distinction. Trying to define a split-point based on protocol data would lead to the problems that participants may –despite clear instructions and practice- not always verbalize

everything that comes to mind, and that there may be a pause in between utterances relating to different processes. In that case, one has the same problem of trying to decide where one process ends and the other begins, only at a smaller scale. Third, the performance efficiency measure we used seems adequate to distinguish between individuals from one participant group that differ significantly on those variables. However, this is a quite relative distinction, and does not position an individual on an exact point of the continuum from novice to expert. Since expertise and instructional design research would benefit from a way to classify students at different sublevels of expertise, it would be interesting to study the potential of this measure at a larger scale. Fourth, it would be interesting to conduct more research on the relationship between the cognitive demands measured through eye movement data and more traditional measures of cognitive load such as self-report or secondary task data. Finally, a replication with larger sample sizes is desirable, because these small sample sizes do not warrant any definite claims.

Nonetheless, the results of this study suggest that the combined use of verbal-protocol and eye-movement data can enhance our insight into (implicit) cognitive processes. Although concurrent verbal protocol data reveal important information on the content of the cognitive processes, eye fixation data provide insights into processing demands and specific content of processes, that is, on the allocation of attention to specific components, which cannot be obtained from verbal protocol data. Therefore, studies using a combination of eye movement and verbal protocol data may contribute to our knowledge of the microstructure of expertise development in a domain and provide valuable input for instructional design.

References

- Alexander, P.A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32, 10-14.
- Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1322-1340.
- Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory and Cognition*, 29, 1146-1152.
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41, 801-849.
- Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition*, 7, 149-158.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence for maximal adaptation to task constraints. *Annual Review of Psychology*, 47, 273-305.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.
- Gitomer, D. H. (1988). Individual differences in technical troubleshooting. *Human Performance*, 1, 111-131.
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *TRENDS In Cognitive Sciences*, 5, 236-243.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 172-190.
- Kalyuga, S., & Sweller, J. (in press). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology, Research and Development*.
- Lankford, C. (2000). GazeTracker™. Software designed to facilitate eye movement analysis.

- Proceedings of the Eye Tracking Research and Applications Symposium* (pp.57-63). New York: ACM Press.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*, 429-434.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.
- Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1994). Diagnostic reasoning and medical expertise. In D. L. Medin (Ed.), *Psychology of learning and motivation: Advances in research and theory* (pp. 187-252). San Diego, CA: Academic Press.
- Patel, V. L., Groen, G. J., & Norman, G. R. (1993). Reasoning and instruction in medical curricula. *Cognition and Instruction, 10*, 335-378.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.
- Rikers, R. M. J. P., Schmidt, H. G., & Boshuizen, H. P. A. (2000). Knowledge encapsulation and the intermediate effect. *Contemporary Educational Psychology, 25*, 150-166.
- Salvucci, D. D. (1999). Mapping eye movements to cognitive processes (Doctoral dissertation, Carnegie Mellon University, 1999). *Dissertation Abstracts International, 60*, 5619.
- Schaafstal, A., Schraagen, J. M., & Van Berlo, M. (2000). Cognitive task analysis and innovation of training: The case of structured troubleshooting. *Human Factors, 42*, 75-86.
- Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education*. Hillsdale, NJ: Erlbaum.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
- Shute, V., & Towle, B. (2003) Adaptive e-learning. *Educational Psychologist, 38*, 105-114.
- Stelmach, L. B., Campsall, J. M., & Herdman, C. M. (1997). Attentional and ocular movements. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 823-844.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.
- Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, *32*, 9-31.
- Tuovinen, J. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science*, *32*, 133-152.
- Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*, 629-646.
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology*, *57*, 165-182.
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*, 167-174.
- Van Gog, T., Ericsson, K. A., Rikers, R. M. J. P., & Paas, F. (in press). Instructional design for advanced learners: Establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *Educational Technology, Research and Development*.
- Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology*, *89*, 231-247.

Author Note

Correspondence concerning this article should be addressed to Tamara van Gog, Open University of the Netherlands, Educational Technology Expertise Center, P.O. Box 2960, 6401 DL, Heerlen, the Netherlands. E-mail: tamara.vangog@ou.nl

The authors gratefully acknowledge the assistance of Puk Witte in collecting the data reported in this article. This research project is funded by the Netherlands Organization for Scientific Research (NWO, The Hague, project no. 411-01-010).

Footnotes

¹ The medical domain is somewhat of an exception; see for example Rikers, Schmidt, & Boshuizen (2000).

² It is generally held that a movement of attention precedes the movement of the eyes, but it must be noted that although this is the case in response to exogenous cues, findings by Stelmach, et al. (1997) suggest that this might not always be so in response to endogenous cues.

³ Had the participants been further in the intermediate phase, one would expect the opposite pattern, as both would likely build representations, but the higher expertise participants would have been faster at this.

⁴ The performance efficiency scores of all participants were calculated by subtracting the mean standardized mental effort score (zE) and the mean standardized time-on-task (zT) from the mean standardized performance score (zP), and dividing the outcome by the square root of 3:

$$\frac{zP - zE - zT}{\sqrt{3}}$$

These mean standardized scores were based on the scores on all tasks in the experiment. The reader is referred to Tuovinen & Paas (2004) for a discussion of this formula in relation to determining the efficiency of instructional conditions.

⁵ This excludes fixations outside the diagram, for example those on the toolbar.

⁶ The same pattern emerges when looking at phase 2 and 3 in their entirety instead of at the subphases: there are no significant differences in the mean fixation duration over phases for the ‘lower expertise’ students ($\chi^2 = .40$, $df = 4$, $p = .954$), and significant differences are found for the ‘higher expertise’ students ($\chi^2 = 7.60$, $df = 4$, $p = .024$), with the mean fixation duration in phase 1 being significantly lower than that in the other phases.

Table 1

Means and Standard Deviations of the Lower and Higher Expertise Participants' Mental Effort, Performance, and Time-on-Task Scores on All Experimental Tasks

	Lower		Higher		Significant (2-tailed)
	M	SD	M	SD	
Mental effort (scale 1-9)	5.28	.42	3.63	.54	$U=0, p=0.008$
Performance (scale 1-7)	4.55	.40	5.69	.34	$U=0, p=0.008$
Time-on-task (seconds)	359.98	45.46	163.29	55.47	$U=0, p=0.008$

Table 2

Means and Standard Deviations of the Lower and Higher Expertise Participants' Absolute Time Spent, Relative Time Spent, Number of Fixations Made, and Mean Fixation Duration per Phase

Phase	Time (s)				Time (%)				Mean Nr. Fixations				Mean Fixation Duration (ms)			
	L		H		L		H		L		H		L		H	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1	7.48	6.70	10.14	5.07	2.55	3.07	7.15	2.94	19.60	20.07	23.40	13.63	255.8	86.5	195.5	17.8
2	14.94	6.31	22.40	21.41	4.40	3.44	17.78	20.68	35.40	24.24	50.80	56.79	261.7	54.9	343.9	85.3
2.1	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	21.60	13.07	24.50	28.52	227.1	27.2	355.5	123.6

2.2	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	13.80	11.63	26.30	28.79	346.2	135.7	370.1	146.6
3	12.66	12.87	15.07	7.30	4.38	5.21	10.87	5.09	23.4	13.83	32.40	22.32	254.3	75.0	308.2	60.6
3.1	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	11.60	8.11	16.00	11.45	285.2	84.7	281.6	68.5
3.2	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	11.60	6.07	16.40	11.44	247.0	90.1	346.2	100.7

Note. L = lower expertise; H = higher expertise; n.a. = not applicable

Table 3

The First Quartile, Median (in Bold), and Third Quartile Values for the Relative Time Spent and Mean Fixation Duration per Phase

Phase	Time (%)		Mean fixation duration (ms)	
	Lower expertise	Higher expertise	Lower expertise	Higher expertise
1	.47 - 1.03* - 5.38	4.49 - 7.33* - 9.70	202.1 - 217.5* - 328.6	181.2 - 201.6* - 206.9
2	1.66 - 2.85 - 7.92	4.39 - 7.34 - 36.40	219.8 - 245.6 - 311.6	271.7 - 317.7 - 429.3
2.1	n.a.	n.a.	204.7 - 224.3* - 250.9	255.2 - 348.4* - 459.5
2.2	n.a.	n.a.	241.7 - 285.0 - 481.3	256.0 - 330.9 - 503.6
3	.60 - 3.21* - 8.76	6.46 - 9.76* - 15.83	194.9 - 233.9 - 323.9	254.0 - 291.2 - 378.0
3.1	n.a.	n.a.	214.2 - 252.0 - 372.8	211.2 - 302.6 - 341.4
3.2	n.a.	n.a.	169.9 - 225.3 - 334.8	264.0 - 327.6 - 437.7

Note. Medians marked with * differ significantly between groups; n.a. = not applicable.

Table 4

Lower and Higher Expertise Participants' Verbalizations per Phase

P	Phase 1	Phase 2	Phase 3
L1	Uh.. well.....	Uhm ... ok let's (incomprehensible)	Uhm.....
L2	Let's see first what's in the circuit ... hm. Try it..	Hm, well, hm. hm yes, the ammeter is right	That lamp here..
L3	I check first whether the lamp works...	The lamp.. none of the three lamps works.. first ... I myself think that the power supply is too low. Check whether that's the case	So now I'm done...right? Uh..is it off now? Yes it's off right? (Experimenter: "yes, the switch is open now"). Hm, that's not good. ... uhm ... I will lower the power to 15
L4	Yes. Ok. I'll try first to see if it works. Probably not.	.. Uhm.. let me think Yes, I think I will here..or wait, I'll first raise the voltage, maybe there is too little	Well, that was too much ... (laughs). Let's see if 12 ... I can always raise it later ..
L5	Ok I'll see first what it does and does not do.	Ok, it does next to nothing. ...-10 ...-1V..I was thinking first ... it needs more voltage ...	Nothing ... again more ..
H1	Ok...hm....Circuit. 3V. So nothing will happen when I close this, voltage is too low.	Let's just raise the voltage a little	Then all three of them burn. Those glow a bit slow, uh, soft. Switch is still open...has to be closed
H2	MISSING	MISSING	MISSING

H3	Well, what do we have here? Just two lamps with all kinds...aha, yes, so this one's wrong...oh no, this is a meter. Uh sorry, resistor. Two lamps. These are in series. In principle they would have to burn when I close this	But the voltage is too low. What do I think...what do I think now? This one has too be raised. 3.5, 4.5 should it be. But oh, there's also a resistor.	Now see what it does. That lamp can only take 9V...and this one to. So when I'd connect them in series ...this one could ...9.5 ... yeah ..could be...they burn...
H4	Ok..uhm.. first close this	It does nothing at all.	That's more like it
H5	Let's see, a switch, so we'll check first what will blow, then we can work towards that. I close it.	And nothing burns. 10mA, that's very strange, but let's see ... uh ..does it say on those lamps how much...oh. 6mA should go through it then ...or am I wrong? Well, no I'm not wrong. Uh ..off... I think the current is too low, so I'll raise the voltage here.	Well, those two already burn a little...this one still does nothing. Because the switch is open of course. Those lamps, how bright should they burn, because..40mA.. 20...

Note. P = participant; L = lower expertise, H = higher expertise, m = milli, A = Ampère, V = Volt

Figure Captions

Figure 1. The troubleshooting task (text added).

Figure 2. Fixation durations by expertise level (lower, higher) over the phases.



